



## Non-contiguous finished genome sequence and description of *Streptococcus varani* sp. nov.

Sofiane Bakour, Jaishriram Rathored, Cheikh Ibrahima Lo, Oleg Mediannikov, Mamadou Beye, Cyrille Bilé Ehounoud, Philippe Biagini, Didier Raoult, Pierre-Edouard Fournier, Florence Fenollar

### ► To cite this version:

Sofiane Bakour, Jaishriram Rathored, Cheikh Ibrahima Lo, Oleg Mediannikov, Mamadou Beye, et al.. Non-contiguous finished genome sequence and description of *Streptococcus varani* sp. nov.. *New Microbes and New Infections*, 2016, 11, pp.93-102. 10.1016/j.nmni.2016.03.004 . hal-01303332

**HAL Id: hal-01303332**

**<https://hal.science/hal-01303332>**

Submitted on 18 Apr 2018

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution - NonCommercial - NoDerivatives 4.0 International License

# Non-contiguous finished genome sequence and description of *Streptococcus varani* sp. nov.

S. Bakour<sup>1</sup>, J. Rathored<sup>1</sup>, C. I. Lo<sup>1,2</sup>, O. Mediannikov<sup>1,2</sup>, M. Beye<sup>1</sup>, C. B. Ehounoud<sup>1</sup>, P. Biagini<sup>3</sup>, D. Raoult<sup>1,2,4</sup>, P.-E. Fournier<sup>1</sup> and F. Fenollar<sup>1,2</sup>

1) Aix-Marseille Université, URMITE, UM63, CNRS 7278, IRD 198, Inserm U1095, Faculté de médecine, Marseille cedex 05, France, 2) Campus international UCAD-IRD, Dakar, Senegal, 3) Equipe Emergence et Co-évolution Virale, UMR 7268 – Aix-Marseille Université/ EFS / CNRS, IHU Méditerranée Infection, Marseille, France and 4) Special Infectious Agents Unit, King Fahd Medical Research Center, King Abdul Aziz University, Jeddah, Saudi Arabia

## Abstract

Strain FF10<sup>T</sup> (= CSUR PI489 = DSM 100884) was isolated from the oral cavity of a lizard (*Varanus niloticus*) in Dakar, Senegal. Here we used a polyphasic study including phenotypic and genomic analyses to describe the strain FF10<sup>T</sup>. Results support strain FF10<sup>T</sup> being a Gram-positive coccus, facultative anaerobic bacterium, catalase-negative, non-motile and non-spore forming. The sequenced genome counts 2.46 Mb with one chromosome but no plasmid. It exhibits a G+C content of 40.4% and contains 2471 protein-coding and 45 RNA genes. On the basis of these data, we propose the creation of *Streptococcus varani* sp. nov.

New Microbes and New Infections © 2016 The Authors. Published by Elsevier Ltd on behalf of European Society of Clinical Microbiology and Infectious Diseases.

**Keywords:** Culturomics, genome, *Streptococcus varani*, taxonogenomics, *Varanus niloticus*

**Original Submission:** 5 February 2016; **Accepted:** 14 March 2016

**Article published online:** 19 March 2016

**Corresponding author:** P.-E. Fournier, URMITE UM63, CNRS7278, IRD198, INSERM1095, Faculté de Médecine, Aix Marseille Université, 27 boulevard Jean Moulin, 13385 Marseille cedex 5, France  
**E-mail:** pierre-edouard.fournier@univ-amu.fr

## Introduction

The genus *Streptococcus* contains 113 recognized species and 22 subspecies (<http://www.bacterio.net/streptococcus.html>) as of 5 February 2016. These species are Gram-positive cocci, chain-forming, facultative anaerobes and catalase negative [1]. Many of them are associated with human or animal hosts. Previous studies have reported that a large number of *Streptococcus* species colonize the oral cavities of humans and animals [2]. Among oral streptococci isolated from animals, we note: *Streptococcus ursoris* isolated from the oral cavities of bears [1], *Streptococcus orisratti* isolated from the surface of the lower

molars of Sprague-Dawley rats [3], *Streptococcus oriloxodontae* isolated from the oral cavities of elephants [4], and *Streptococcus mutans* isolated from dental plaque [5].

Recently, with next-generation sequencing technology able to sequence whole genomes in a short time, and mass spectrometric analysis of bacteria, we have had easy access to genetic and proteomic information [6]. Therefore, we propose a polyphasic approach combining genomic properties in combination with matrix-assisted laser-desorption/ionization time-of-flight (MALDI-TOF) spectra and phenotypic characteristics to describe new bacterial species. The strain FF10<sup>T</sup> (= CSUR PI489 = DSM 100884) was isolated from the oral cavity of the Nile monitor (*Varanus niloticus*) in Dakar, Senegal.

Here, we present a summary classification and a set of features for *Streptococcus varani* sp. nov. strain FF10<sup>T</sup> (= CSUR PI489 = DSM 100884), including the description of its complete genome and annotation. These characteristics support the circumscription of the species *S. varani*.

## Classification and Features of the Strain

### Strain identification

In 2014, a sample was collected from the oral cavity of a lizard reptile named 'Nile monitor' (*Varanus niloticus*) in Dakar, Senegal, and stored at  $-80^{\circ}\text{C}$ . In October 2014, the strain FF10<sup>T</sup> (Table 1) was isolated from this sample by cultivation on 5% sheep blood-enriched Columbia agar (BioMérieux, Marcy l'Etoile, France) at  $37^{\circ}\text{C}$  with 5%  $\text{CO}_2$ .

To identify the strain, MALDI-TOF MS protein analysis was performed as previously described [7] using a Microflex spectrometer (Bruker Daltonics, Leipzig, Germany). The 12 FF10<sup>T</sup> spectra were imported into the MALDI BioTyper software (version 2.0, Bruker) and analysed by standard pattern matching (with default parameter settings) against the main spectra of 6252 bacteria. From the resulting scores, the tested species may or may not be identified compared with the instrument's database; a score  $\geq 2$  with a validly published species enables identification at the species level; a score  $\geq 1.7$  and  $< 2$  allows identification at the genus level; and a score  $< 1.7$  does not enable any identification. For strain FF10<sup>T</sup>, the scores obtained were lower than 1.306, suggesting that our strain was not a member of any known species. The reference mass spectrum of *S. varani* strain FF10<sup>T</sup> was added to our database (Fig. 1). A gel view comparing the spectrum of strain FF10<sup>T</sup> with those of other *Streptococcaceae* species is shown in Fig. 2. The bacterium was identified using 16S rRNA PCR coupled with sequencing, as previously described [8]. Strain FF10<sup>T</sup> exhibited 96% 16S rRNA sequence similarity with *Streptococcus minor* strain DSM 17118

(GenBank accession number: AY232832), the phylogenetically closest bacterial species with standing in nomenclature-validated *Streptococcus* species (Fig. 3). This value was lower than the 98.7% 16S rRNA gene sequence threshold recommended by Stackebrandt and Ebers to delineate a new species without carrying out DNA–DNA hybridization [9]. *Streptococcus varani* strain FF10<sup>T</sup> 16S rRNA accession number from GenBank Sequence Database is LN810501.

### Growth characterization

Different growth temperatures ( $25^{\circ}\text{C}$ ,  $30^{\circ}\text{C}$ ,  $37^{\circ}\text{C}$ ,  $45^{\circ}\text{C}$  and  $56^{\circ}\text{C}$ ) were tested. Growth was obtained between 25 and  $37^{\circ}\text{C}$ , with optimal growth at  $37^{\circ}\text{C}$ . Growth of the strain was tested also under anaerobic and microaerophilic conditions using GENbag anaer and GENbag microaer systems, respectively (bioMérieux), and under aerobic conditions, with or without 5%  $\text{CO}_2$ . Strain growth was observed under anaerobic and microaerophilic conditions but optimal growth was observed under aerobic conditions. Colonies were translucent and yellow with a regular surface, haemolytic on 5% sheep blood-enriched Columbia agar (bioMérieux), and approximately 1 mm in diameter. A motility test was negative. Cells were Gram-positive cocci, unable to form spores (Fig. 4), and with mean diameter of  $0.6\text{ }\mu\text{m}$  (range  $0.4\text{--}0.8\text{ }\mu\text{m}$ ) and mean length of  $1.3\text{ }\mu\text{m}$  (range  $0.7\text{--}1.9\text{ }\mu\text{m}$ ) (Fig. 5).

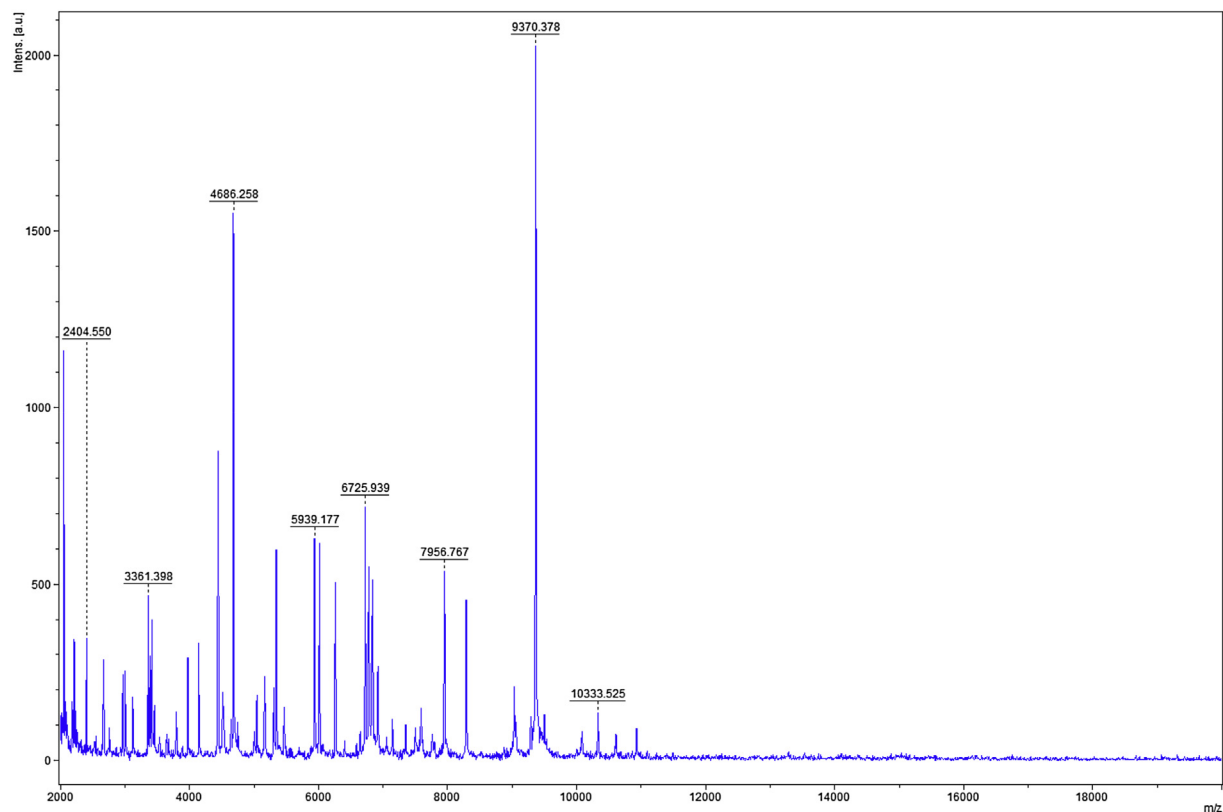
### Biochemical characterization and antibiotic susceptibility

This bacterium, FF10<sup>T</sup>, exhibits neither catalase nor oxidase activities. Using an API ZYM strip (bioMérieux), positive reactions were observed for alkaline phosphatase, esterase, esterase-lipase, lipase, leucine arylamidase, acid phosphatase,  $\beta$ -glucosidase,  $\beta$ -galactosidase,  $\alpha$ -mannosidase, and  $\alpha$ -fucosidase. Negative reactions were noted for cystine arylamidase, valine arylamidase, trypsin,  $\alpha$ -chymotrypsin,  $\alpha$ -glucosidase,  $\alpha$ -galactosidase,  $\beta$ -glucuronidase, naphthol-AS-BI-phosphohydrolase, and *N*-acetyl- $\beta$ -glucosaminidase. Using an API 20 NE strip (bioMérieux), positive reactions were observed only for the esculin hydrolysis test whereas negative reactions were observed for nitrate reduction, urease, indole production, arginine dihydrolase, glucose fermentation, arabinose, mannose, mannitol, *N*-acetyl-glucosamine, maltose, gluconate, caprate, adipate, malate, citrate, phenyl-acetate assimilation, and gelatine hydrolysis.

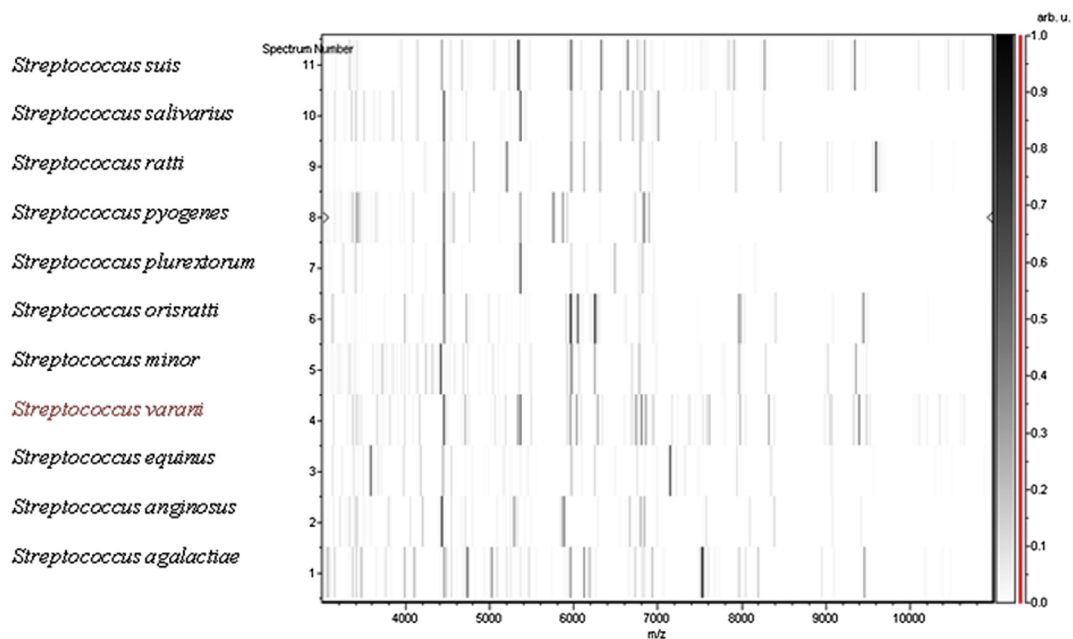
Using API 50 CH strip (bioMérieux), negative reactions were observed for the fermentation of glycerol, erythritol, D-arabinose, L-arabinose, D-ribose, D-xylose, L-xylose, D-adonitol, methyl- $\beta$ -D-xylopyranoside, D-galactose, D-glucose, D-fructose, D-mannose, L-sorbose, L-rhamnose, dulcitol, inositol, D-mannitol, D-sorbitol, methyl- $\alpha$ -D-xylopyranoside, methyl- $\alpha$ -D-glucopyranoside, *N*-acetylglucosamine, amygdalin, arbutin,

**TABLE 1.** Classification and general features of *Streptococcus varani* strain FF10<sup>T</sup>

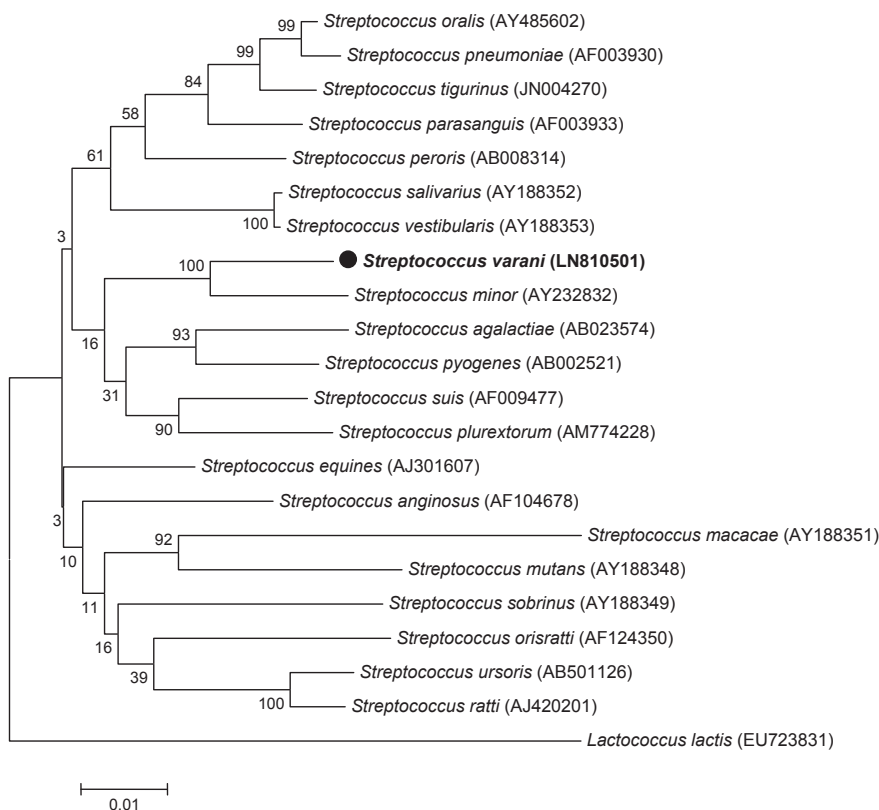
Property	Term	References
Classification	Domain <i>Bacteria</i> Phylum <i>Firmicutes</i> Class <i>Bacilli</i> Order <i>Lactobacillales</i> Family <i>Streptococcaceae</i> Genus <i>Streptococcus</i> Species <i>Streptococcus varani</i> Type strain FF10 <sup>T</sup>	[25] [26,27] [28,29] [29,30] [31,32] [31,33,34]
Gram stain	Positive	
Cell shape	Cocci	
Motility	Non-motile	
Sporulation	Non-spore forming	
Temperature range	Mesophile	
Optimum temperature	$37^{\circ}\text{C}$	
Carbon source	Unknown	
Habitat	Lizard	
Salinity	Unknown	
Oxygen requirement	Facultative anaerobe	
Biotic relationship	Free living	
Pathogenicity	Unknown	
Geographic location	Dakar	
Sample collection	October 22, 2014	
Latitude	13.7167	
Longitude	$-16.4167$	
Altitude	51 m above sea level	



**FIG. 1.** Reference mass spectrum from *Streptococcus varani* sp. nov. strain FF10<sup>T</sup>. Spectra from 12 individual colonies were compared and a reference spectrum was generated.



**FIG. 2.** Gel view comparing *Streptococcus varani* sp. nov. strain FF10<sup>T</sup> spectra with other members of the *Streptococcus* genus. The gel view displays the raw spectra of loaded spectrum files arranged in a pseudo-gel like look. The x-axis records the m/z value. The left y-axis displays the running spectrum number originating from subsequent spectra loading. The peak intensity is expressed by a Grey-scale scheme code. The colour bar and the right y-axis indicate the relation between the colour in which a peak is displayed and the peak intensity in arbitrary units. Displayed species are indicated on the left.



**FIG. 3.** Phylogenetic tree showing the position of *Streptococcus varani* sp. nov. strain FF10<sup>T</sup> relative to other type strains within the genus *Streptococcus*. The strains and their corresponding GenBank accession numbers for 16S rRNA genes are in parenthesis. The tree was constructed by the neighbour joining method with Kimura two-parameter model and 1000 bootstrap replications using the MEGA6 software and rooted by using *Lactococcus lactis* (EU723831) strain as the out-group.

salicin, D-cellobiose, D-maltose, D-lactose, D-melibiose, D-saccharose, D-trehalose, inulin, D-melezitose, D-raffinose, amidon, glycogen, xylitol, gentiobiose, D-turanose, D-lyxose, D-tagatose, D-fucose, L-fucose, D-arabitol, L-arabitol, potassium gluconate, potassium 2-ketogluconate and potassium-5-ketogluconate.

*Streptococcus varani* strain FF10<sup>T</sup> was susceptible to penicillin, amoxicillin, amoxicillin/clavulanic acid, ticarcillin, ceftriaxone, cefalotin, imipenem, gentamicin, kanamycin, trimethoprim/sulfamethoxazole, erythromycin, doxycycline, metronidazole,

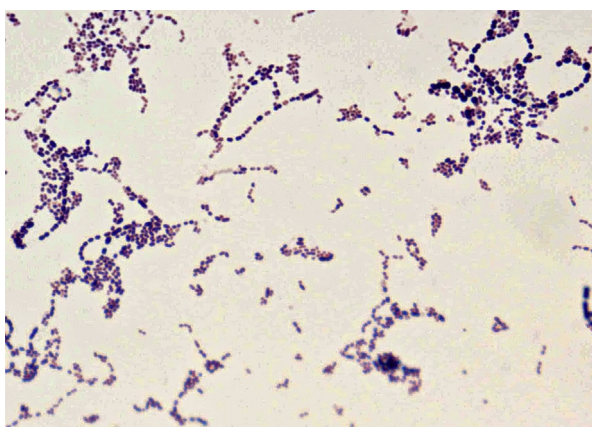
vancomycin and rifampicin, but resistant to ciprofloxacin, nitrofurantoin and colistin.

The differential phenotypic characteristics with other *Streptococcus* species (*S. minor* [10], *Streptococcus plurextorum* [11], *S. ursoris* [1] and *S. orisratti* [3]) are summarized in Table 2.

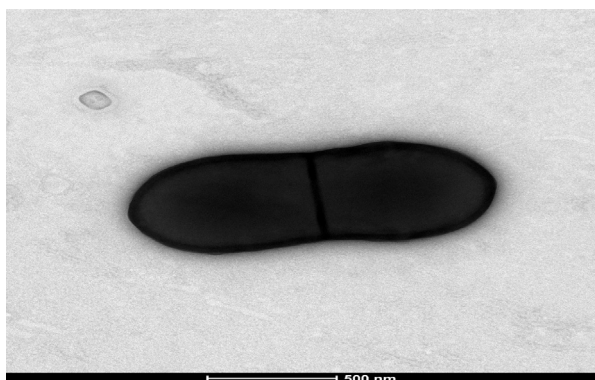
## Genome Description

### Genome sequencing and assembly

The DNA was extracted using the phenol/chloroform method as previously described [12]. Genomic DNA (gDNA) of *S. varani* FF10<sup>T</sup> was sequenced on the MiSeq Technology (Illumina Inc., San Diego, CA, USA) using the mate-pair strategy. The gDNA was bar-coded in order to be mixed with 11 other projects with the Nextera Mate Pair sample prep kit (Illumina). The gDNA was quantified by a Qubit assay with the high sensitivity kit (Thermo Fisher Scientific, Waltham, MA, USA) to 62.4 mg/L. The mate-pair library was prepared with 1.5 µg of genomic DNA using the Nextera mate-pair Illumina guide. The gDNA sample was simultaneously fragmented and tagged with a mate-pair junction adapter. The fragmentation pattern was validated on an Agilent 2100 BioAnalyzer (Agilent Technologies Inc., Santa Clara, CA, USA) with a DNA 7500 LabChip. The



**FIG. 4.** Gram staining of *Streptococcus varani* sp. nov. strain FF10<sup>T</sup>.



**FIG. 5.** Transmission electron microscopy of *Streptococcus varani* sp. nov. strain FF10<sup>T</sup>, using a TECNAI G20 (FEI) at an operating voltage of 200 keV. The scale bar represents 500 nm.

DNA fragments ranged from 1.5 kb up to 11 kb with optimal size of 5.8 kb. No size selection was performed and 591 ng of tagged fragments were circularized. The circularized DNA was mechanically sheared to small fragments, optimally at 697 bp, on the Covaris device S2 in T6 tubes (Covaris, Woburn, MA, USA). The library profile was visualized on a High Sensitivity Bioanalyzer LabChip (Agilent Technologies Inc.) and the final concentration library was measured at 75.52 nmol/

L. The libraries were normalized at 2 nM and pooled. After a denaturation step and dilution at 15 pM, the pool of libraries was loaded onto the reagent cartridge and then onto the instrument along with the flow cell. Automated cluster generation and sequencing were performed in a single 39-h run in a 2 × 251-bp.

Total information of 5.9 Gb was obtained from a 624 K/mm<sup>2</sup> cluster density with a cluster passing quality control filters of 96.33% (12 040 000 clusters). Within this run, the index representation for *S. varani* FF10<sup>T</sup> was determined to 5.26%. The 610 160 paired reads were filtered according to the read qualities. These reads were trimmed then assembled.

## Genome Annotation and Genome Analysis

Open reading frames (ORFs) were predicted using PRODIGAL (<http://prodigal.ornl.gov/>) with default parameters. However, the predicted ORFs were excluded if they spanned a sequencing gap region. The predicted bacterial protein sequences were searched against GenBank [13] and Clusters of Orthologous Groups (COG) databases using BLASTP. The tRNAs and rRNAs were predicted using the tRNAscan-SE [14] and RNAmmer [15] tools, respectively. Signal peptides and numbers of transmembrane helices were predicted using

**TABLE 2.** Differential characteristics of *Streptococcus varani* sp. nov. strain FF10<sup>T</sup> with other strains (*Streptococcus minor* [8], *Streptococcus plurextorum* [9], *Streptococcus ursoris* [1] and *Streptococcus orisratti* [3])

Characteristics	<i>S. varani</i>	<i>S. minor</i>	<i>S. plurextorum</i>	<i>S. ursoris</i>	<i>S. orisratti</i>
Cell diameter (µm)	0.4–0.8	<1	na	0.5–0.7	na
Oxygen requirement	Facultative anaerobic	Microaerobic	Facultative anaerobic	Facultative anaerobic	Facultative anaerobic
Gram stain	+	+	+	+	+
Motility	–	na	–	na	–
Endospore formation	–	na	na	–	na
Production of:					
Alkaline phosphatase	+	–	–	–	–
Acid phosphatase	+	+	+	+	+
Catalase	–	na	–	+	–
Oxidase	–	na	–	na	na
Nitrate reductase	–	na	na	na	na
Urease	–	–	–	–	na
α-galactosidase	–	W	+	+	–
β-galactosidase	+	–	+	–	–
β-glucuronidase	–	–	+	–	–
α-glucosidase	–	na	–	+	na
β-glucosidase	+	na	–	+	na
Esterase	+	na	–	na	na
Esterase lipase	+	na	–	na	na
Naphthol-AS-BI-phosphohydrolase	–	na	–	+	+
N-acetyl-β-glucosaminidase	–	na	–	–	na
Utilization of:					
Gluconate	–	–	DP	na	na
D-ribose	–	–	–	+	–
D-xylose	–	–	–	–	na
D-fructose	–	+	+	+	+
D-glucose	–	+	+	+	+
D-mannose	–	+	–	+	+
D-lactose	–	+	+	+	+
D-maltose	–	+	+	+	+
DL-arabinose	–	–	–	–	–
Habitat	Lizard	Dog, cat and calf	Pigs	Bears	Rats

+, positive; –, negative; DP, delayed production; W, weak reaction; na, not available.



**TABLE 3.** Nucleotide content and gene count levels of the genome

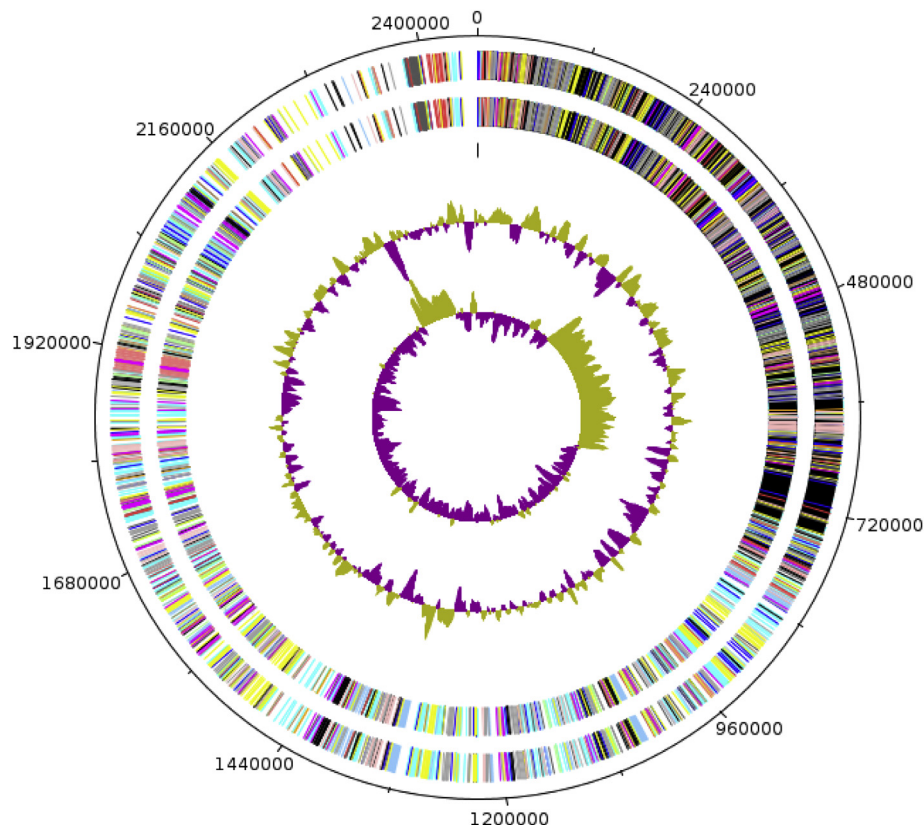
Attribute	Value	% of total <sup>a</sup>
Genome size (bp)	2 460 376	100
DNA coding (bp)	2 170 614	88.2
DNA G+C (bp)	993 991	40.4
Total genes	2516	100
Protein coding genes	2471	98.21
RNA genes	45	1.78
Genes with function prediction	1852	73.60
Genes assigned to COGs	1746	69.39
Genes with signal peptides	118	4.68
Genes with transmembrane helices	565	22.45
Genes associated with resistant genes	0	0
Genes associated with PKS or NRPS	5	0.19
ORFans genes	0	0
CRISPRs repeats	0	0
Genes associated with Pfam-A domains	2264	89.98

<sup>a</sup>The total is based on either the size of the genome in base pairs or the total number of protein coding genes in the annotated genome.

SignalP [16], and TMHMM [17], respectively. Mobile genetic elements were predicted using PHAST [14] and RAST [18]. ORFans were identified if their BLASTP *E*-value was lower than 1e-03 for alignment length >80 amino acids. If alignment lengths were <80 amino acids, we used an *E*-value of 1e-05.

Such parameter thresholds have already been used in previous work to define ORFans. ARTEMIS [19] and DNA PLOTTER [20] were used for data management and visualization of genomic features, respectively. The MAUVE alignment tool (version 2.3.1) was used for multiple genomic sequence alignment [21].

The mean level of nucleotide sequence similarity at the genome level between *S. varani* and other bacteria (*S. orisratti* (ARCG01000001.1), *S. tigurinus* (AORU01000001.1), *Streptococcus parasanguinis* (CP002843.1), *Streptococcus agalactiae* (AEQQ01000001.1), *S. plurextorum* (AUIO01000001.1), *S. minor* (AQYB01000001.1), *Streptococcus oralis* (ADMV01000001.1), *Streptococcus pyogenes* (AE004092.2), *Streptococcus vestibularis* (AEV101000001.1), *Streptococcus peroris* (AEVF01000001.1) and *Lactococcus lactis* (NC\_002662.1)) was estimated using the Average Genomic Identity of gene Sequences (AGIOS) home-made software [22]. Overall, this software combines the functionality of other software programs: PROTEINORTHO [23] (detects orthologous proteins between genomes compared two by two, then retrieves the corresponding genes) and the Needleman–Wunsch global



**FIG. 6.** Graphical circular map of the chromosome. From outside to the centre. Genes on the forward strand coloured by COG categories (only genes assigned to COG), genes on the reverse strand coloured by COG categories (only gene assigned to COG), RNA genes (tRNAs green, rRNAs red), GC content, and GC skew.

**TABLE 4.** Number of genes associated with the 25 general COG functional categories<sup>a</sup>

Code	Value	%	Description
J	144	8.18	Translation
A	0	0	RNA processing and modification
K	101	5.74	Transcription
L	91	5.17	Replication, recombination and repair
B	0	0	Chromatin structure and dynamics
D	23	1.31	Cell cycle control, mitosis and meiosis
Y	0	0	Nuclear structure
V	37	2.10	Defence mechanisms
T	41	2.33	Signal transduction mechanisms
M	83	4.71	Cell wall/membrane biogenesis
N	5	0.28	Cell motility
Z	0	0	Cytoskeleton
W	0	0	Extracellular structures
U	20	1.13	Intracellular trafficking and secretion
O	54	3.07	Post-translational modification, protein turnover, chaperones
C	37	2.10	Energy production and conversion
G	111	6.30	Carbohydrate transport and metabolism
E	145	8.24	Amino acid transport and metabolism
F	69	3.92	Nucleotide transport and metabolism
H	42	2.39	Coenzyme transport and metabolism
I	35	1.99	Lipid transport and metabolism
P	71	4.03	Inorganic ion transport and metabolism
Q	13	0.74	Secondary metabolites biosynthesis, transport and catabolism
R	156	8.86	General function prediction only
S	154	8.75	Function unknown
—	328	18.64	Not in COGs

<sup>a</sup>The total is based on the total number of protein-coding genes in the annotated genome.

alignment algorithm (determines the mean percentage of nucleotide sequence identity among orthologous ORFs).

### Genome properties

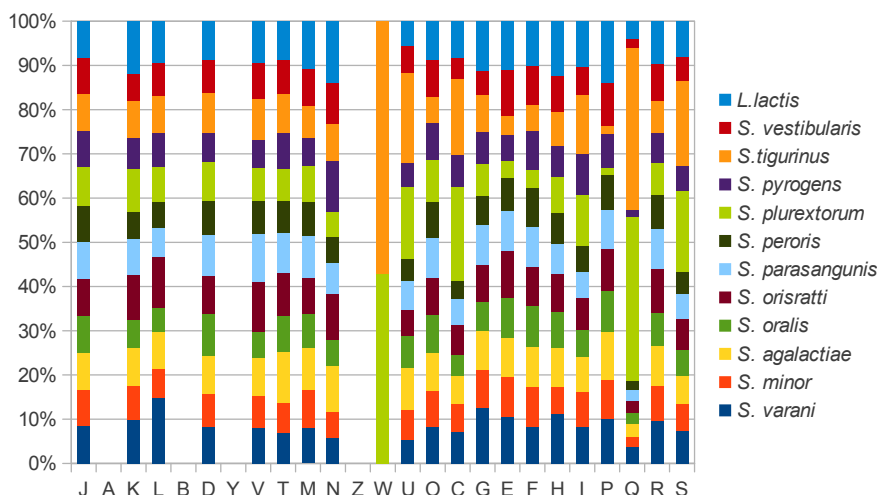
The genome is 2 460 376 bp long with 40.4% GC content (Table 3). It is composed of 16 contigs (Fig. 6). Of the 2516 predicted genes, 2471 were protein-coding genes, and 45 were RNAs (one gene is 5S rRNA, one gene is 16S rRNA, one gene is 23S rRNA, 42 genes are tRNA genes). No genes were identified as ORFans (0.0%). The remaining genes were annotated as

hypothetical proteins (293 genes  $\geq 11.64\%$ ). Table 4 distributes the genes into COG functional categories. The genome sequence has been deposited in GenBank under accession number [GCA\\_001375655](https://www.ncbi.nlm.nih.gov/nuclseq/GCA_001375655).

### Comparison with other genomes

The draft genome of *S. varani* (2.46 Mb) is larger than those of *S. agalactiae* (2.11 Mb), *S. minor* (1.93 Mb), *S. oralis* (1.90 Mb), *S. orisratti* (2.41 Mb), *S. parasanguinis* (2.15 Mb), *S. peroris* (1.63 Mb), *S. plurextorum* (2.10 Mb), *S. pyogenes* (1.85 Mb), *S. tigurinus* (2.18 Mb), *S. vestibularis* (1.84 Mb). The G+C content of *S. varani* (40.40%) is lower than those of *S. minor* (41.10%), *S. oralis* (44.40%), *S. parasanguinis* (41.70%), *S. plurextorum* (41.10) but higher than those of *S. agalactiae* (35.30%), *S. orisratti* (38.50%), *S. peroris* (39.10%), *S. pyogenes* (38.50%), *S. tigurinus* (40.30%) and *S. vestibularis* (39.60%).

Similarly, the protein-coding genes of *S. varani* (2471) are larger than those of *S. agalactiae*, *S. minor*, *S. oralis*, *S. orisratti*, *S. parasanguinis*, *S. peroris*, *S. plurextorum*, *S. pyogenes*, *S. tigurinus* and *S. vestibularis* (2139, 1903, 1795, 2319, 1982, 1608, 2081, 1776, 2146 and 1812, respectively). The gene content of *S. varani* (2516) is larger than that of *S. agalactiae*, *S. minor*, *S. oralis*, *S. orisratti*, *S. parasanguinis*, *S. peroris*, *S. plurextorum*, *S. pyogenes*, *S. tigurinus* and *S. vestibularis* (2164, 1948, 1830, 2348, 2055, 1641, 2116, 1854, 2197 and 1838, respectively). However, the COG category gene distribution is similar in all genomes compared (Fig. 7). Among species with standing in nomenclature, AGIOS values ranged from 81.54 between *S. peroris* and *S. oralis* to 65.60 between *L. lactis* and *S. varani* (Table 5). To evaluate the genomic similarity among the strains, we determined two parameters, dDDH, which exhibits high correlation with DDH [24], and AGIOS [22], which was designed to be independent of DDH (Table 6).

**FIG. 7.** Distribution of predicted genes of *Streptococcus varani* strain FF10<sup>T</sup> and 11 other closely related species into COG categories.



**TABLE 5.** The numbers of orthologous protein shared between genomes (upper right)<sup>a</sup>

	<i>S. varani</i>	<i>S. minor</i>	<i>S. agalactiae</i>	<i>S. oralis</i>	<i>S. orisratti</i>	<i>S. parasanguinis</i>	<i>S. peroris</i>	<i>S. plurextorum</i>	<i>S. pyogenes</i>	<i>S. tigurinus</i>	<i>S. vestibularis</i>	<i>L. lactis</i>
<i>Streptococcus varani</i>	<b>2471</b>	476	1079	1016	1133	1067	975	1044	972	1089	983	952
<i>Streptococcus minor</i>	76.99	<b>1903</b>	997	1000	1089	1045	962	1024	930	1023	989	902
<i>Streptococcus agalactiae</i>	69.62	69.74	<b>2139</b>	982	1082	1030	936	1020	1022	1036	1008	915
<i>Streptococcus oralis</i>	70.91	70.91	69.62	<b>1795</b>	1035	1149	1171	994	911	1261	1031	887
<i>Streptococcus orisratti</i>	71.00	70.83	73.03	70.46	<b>2319</b>	1110	1001	1059	976	1092	1066	930
<i>Streptococcus parasanguinis</i>	70.27	70.37	69.26	74.95	69.97	<b>1982</b>	1101	1061	959	1156	1068	915
<i>Streptococcus peroris</i>	70.79	70.69	70.31	81.54	70.73	75.43	<b>1608</b>	950	870	1170	978	845
<i>Streptococcus plurextorum</i>	70.19	70.31	70.95	70.04	72.20	69.49	70.12	<b>2081</b>	951	1031	966	856
<i>Streptococcus pyogenes</i>	70.19	70.00	74.23	70.10	73.04	69.97	70.59	71.26	<b>1176</b>	932	903	840
<i>Streptococcus tigurinus</i>	70.58	70.52	69.82	93.01	70.27	74.91	81.42	69.96	69.92	<b>2146</b>	1051	892
<i>Streptococcus vestibularis</i>	70.47	70.45	71.67	71.82	72.94	72.11	72.00	71.56	72.05	71.69	<b>1812</b>	892
<i>Lactococcus lactis</i>	65.60	65.36	66.90	66.33	66.46	66.03	66.70	65.63	66.68	66.09	66.63	<b>2650</b>

<sup>a</sup>Average percentage similarity of nucleotides corresponding to orthologous protein shared between genomes (lower left) and numbers of proteins per genome (bold).

**TABLE 6.** Pairwise comparison of *Streptococcus varani* with eight other species using GGDC, formula 2 (DDH estimates based on identities / HSP length)<sup>a</sup>

	<i>S. varani</i>	<i>S. minor</i>	<i>S. agalactiae</i>	<i>S. oralis</i>	<i>S. orisratti</i>	<i>S. parasanguis</i>	<i>S. peroris</i>	<i>S. plurextorum</i>	<i>S. pyogenes</i>	<i>S. tigurinus</i>	<i>S. vestibularis</i>
<i>Streptococcus varani</i>	100% ± 0.0	24.7% ± 2.57	25.7% ± 2.53	25.0% ± 2.57	22.4% ± 2.60	25.8% ± 2.56	22.7% ± 2.57	25.2% ± 2.60	26.3% ± 2.72	25.3% ± 2.59	23.1% ± 2.59
<i>Streptococcus minor</i>		100% ± 0.0	23.4% ± 2.53	25.0% ± 2.57	23.2% ± 2.58	24.8% ± 2.73	25.0% ± 3.01	22.7% ± 2.58	27.4% ± 2.59	49.8% ± 3.14	25.4% ± 2.64
<i>Streptococcus agalactiae</i>			100% ± 0.0	25.0% ± 2.57	22.4% ± 2.60	25.8% ± 2.56	22.7% ± 2.57	25.2% ± 2.60	26.3% ± 2.72	25.3% ± 2.59	23.1% ± 2.59
<i>Streptococcus oralis</i>				100% ± 0.0	23.2% ± 2.58	24.8% ± 2.73	25.0% ± 3.01	22.7% ± 2.58	27.4% ± 2.59	49.8% ± 3.14	25.4% ± 2.64
<i>Streptococcus orisratti</i>					100% ± 0.0	24.9% ± 2.57	23.5% ± 2.59	25% ± 2.61	23.2% ± 2.61	22.6% ± 2.58	23.5% ± 2.60
<i>Streptococcus parasanguis</i>						100% ± 0.0	26.3% ± 2.77	23.1% ± 2.58	27.6% ± 2.59	24.5% ± 2.72	27.4% ± 2.68
<i>Streptococcus peroris</i>							100% ± 0.0	24.9% ± 2.57	25.8% ± 2.59	24.9% ± 3.01	25.9% ± 2.65
<i>Streptococcus plurextorum</i>								100% ± 0.0	23.8% ± 2.60	21.9% ± 2.58	23.3% ± 2.58
<i>Streptococcus pyogenes</i>									100% ± 0.0	25.9% ± 2.59	24.7% ± 2.61
<i>Streptococcus tigurinus</i>										100% ± 0.0	26% ± 2.64
<i>Streptococcus vestibularis</i>											100% ± 0.0

<sup>a</sup>The confidence intervals indicate the inherent uncertainty in estimating DDH values from intergenomic distances based on models derived from empirical test data sets (which are always limited in size). These results are in accordance with the 16S rRNA (Fig. 3) and phylogenomic analyses as well as the GGDC results.

## Conclusion

On the basis of phenotypic, phylogenetic and genomic analyses, we formally propose the creation of *Streptococcus varani* sp. nov. that contains the strain FF10<sup>T</sup>. This bacterial strain has been isolated from the oral cavities of a lizard (*Varanus niloticus*) in Dakar, Senegal.

## Description of *Streptococcus varani* sp. nov. strain FF10<sup>T</sup>

*Streptococcus varani* (va.ra'ni. NL. gen. n. *varani*, from *Varanus*, the Latin name from the lizard which the type strain was isolated) is a non-motile, non-spore forming, facultative anaerobe, and Gram-positive coccus. Growth is obtained between 25°C and 37°C, with optimal growth at 37°C. The colonies are translucent and yellow with a regular surface,

haemolytic on 5% sheep blood-enriched Columbia agar, and approximately 1 mm in diameter. Cells have mean diameter of 0.6 µm and mean length of 1.3 µm. *Streptococcus varani* strain FF10<sup>T</sup> exhibits positive reactions for alkaline phosphatase, esterase, esterase-lipase, lipase, leucine arylamidase, acid phosphatase, β-glucosidase, β-galactosidase, α-mannosidase, α-fucosidase, and for the esculin hydrolysis test. *Streptococcus varani* strain FF10<sup>T</sup> is sensitive to penicillin, amoxicillin, amoxicillin/clavulanic acid, ticarcillin, ceftriaxone, cefalotin, imipenem, gentamicin, kanamycin, trimethoprim/sulfamethoxazole, erythromycin, doxycycline, metronidazole, vancomycin and rifampicin, but resistant to ciprofloxacin, nitrofurantoin and colistin. The G+C content of the genome is 40.4%. The 16S rRNA gene sequence and whole-genome shotgun sequence of strain FF10<sup>T</sup> are deposited in GenBank under accession numbers LN810501 and [GCA\\_001375655](https://genbank.ncbi.nlm.nih.gov/GenBank/FASTA/001375655), respectively. The type strain FF10<sup>T</sup> (= CSUR PI489 = DSM 100884) was isolated from the oral cavity of a lizard (*Varanus niloticus*) in Dakar, Senegal.

## Authors' Contributions

SB, JR, and MB performed the genomic analyses and drafted the manuscript. CIL and CBE performed the phenotypic characterization of the bacterium and drafted the manuscript. OM participated in its design and helped to draft the manuscript. PB helped to draft the manuscript. DR conceived the study and helped to draft the manuscript. PEF and FF conceived the study, participated in its design and coordination and helped to draft the manuscript. All authors read and approved the final manuscript.

## Transparency Declaration

The authors declare no conflict of interest.

## Acknowledgements

The authors thank the Xegen Company ([www.xegen.fr](http://www.xegen.fr)) for automating the genomic annotation process. This study was funded by the *Fondation Méditerranée Infection*. We also thank Claudia Andrieu for administrative assistance.

## References

- [1] Shinozaki-Kuwahara N, Takada K, Hirasawa M. *Streptococcus ursoris* sp. nov., isolated from the oral cavities of bears. *Int J Syst Evol Microbiol* 2015;61:40–4.
- [2] Facklam R. What happened to the streptococci: overview of taxonomic and nomenclature changes. *Clin Microbiol Rev* 2002;15:613–30.
- [3] Zhu H, Willcox MDP, Knox KW. A new species of oral *Streptococcus* isolated from Sprague-Dawley rats, *Streptococcus orisratti* sp. nov. *Int J Syst Evol Microbiol* 2000;50:55–61.
- [4] Shinozaki-Kuwahara N, Saito M, Hirasawa M, Takada K. *Streptococcus oriloxodontae* sp. nov., isolated from the oral cavities of elephants. *Int J Syst Evol Microbiol* 2014;64:3755–9.
- [5] Hamada S, Slade H. Biology, immunology, and cariogenicity of *Streptococcus mutans*. *Microbiol Rev* 1980;44:331–84.
- [6] Welker M, Moore ER. Applications of whole-cell matrix-assisted laser desorption/ionization time-of-flight mass spectrometry in systematic microbiology. *Syst Appl Microbiol* 2011;34:2–11.
- [7] Seng P, Drancourt M, Gouriet F, La Scola B, Fournier PE, Rolain JM, et al. Ongoing revolution in bacteriology: routine identification of bacteria by matrix-assisted laser desorption/ionization time-of-flight mass spectrometry. *Clin Infect Dis* 2009;49:543–51.
- [8] Mourembou G, Rathored J, Lekana-Douki JB, Ndjoyi-Mbiguino A, Fenollar F, Michelle C, et al. Non-contiguous finished genome sequence and description of *Kallipyga gabonensis* sp. nov. *New Microbes New Infect* 2015;9:15–23.
- [9] Stackebrandt E, Ebers J. Taxonomic parameters revisited: tarnished gold standards. *Microbiol Today* 2006;33:152–5.
- [10] Vancanneyt M, Devriese L, Haesebrouck F, Baele M, Lefebvre K, Snauwaert C, et al. *Streptococcus minor* sp. nov., from faecal samples and tonsils of domestic animals. *Int J Syst Evol Microbiol* 2004;54:449–52.
- [11] Vela A, Casamayor A, Sanchez Del Rey V, Dominguez L, Fernandez-Garayzabal J. *Streptococcus plurextorum* sp. nov., isolated from pigs. *Int J Syst Evol Microbiol* 2009;59:504–8.
- [12] Lagier JC, Bibi F, Ramasamy D, Azhar EI, Robert C, Yasir M, et al. Non contiguous-finished genome sequence and description of *Clostridium jeddahense* sp. nov.. BASE, EBSCOhost; 2014 (accessed November 24, 2015).
- [13] Benson DA, Karsch-Mizrachi I, Clark K, Lipman DJ, Ostell J, Sayers EW. GenBank. *Nucl Acids Res* 2012;40:D48–53.
- [14] Lowe TM, Eddy SR. tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. *Nucl Acids Res* 1997;25:955–64.
- [15] Lagesen K, Hallin P, Rodland EA, Staerfeldt HH, Rognes T, Ussery DW. RNAmmer: consistent and rapid annotation of ribosomal RNA genes. *Nucl Acids Res* 2007;35:3100–8.
- [16] Bendtsen JD, Nielsen H, von Heijne G, Brunak S. Improved prediction of signal peptides: SignalP 3.0. *J Mol Biol* 2004;340:783–95.
- [17] Krogh A, Larsson B, von Heijne G, Sonnhammer EL. Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. *J Mol Biol* 2001;305:567–80.
- [18] Zhou Y, Liang Y, Lynch KH, Dennis JJ, Wishart DS. PHAST: a fast phage search tool. *Nucl Acids Res* 2011;39:W347–52.
- [19] Rutherford K, Parkhill J, Crook J, Horsnell T, Rice P, Rajandream MA, et al. Artemis: sequence visualization and annotation. *Bioinformatics* 2000;16:944–5.
- [20] Carver T, Thomson N, Bleasby A, Berriman M, Parkhill J. DNAPlotter: circular and linear interactive genome visualization. *Bioinformatics* 2009;25:119–20.
- [21] Darling AC, Mau B, Blattner FR, Perna NT. Mauve: multiple alignment of conserved genomic sequence with rearrangements. *Genome Res* 2004;14:1394–403.
- [22] Ramasamy D, Mishra AK, Lagier JC, Padhmanabhan R, Rossi-Tamisier M, Sentausa E, et al. A polyphasic strategy incorporating genomic data for the taxonomic description of new bacterial species. *Int J Syst Evol Microbiol* 2014;64:384–91.
- [23] Lechner M, Findeiss S, Steiner L, Marz M, Stadler PF, Prohaska SJ. Proteinortho: detection of (co)orthologs in large-scale analysis. *BMC Bioinformatics* 2011;12:124.
- [24] Auch AF, Von Jan M, Klenk HP, Göker M. Digital DNA-DNA hybridization for microbial species delineation by means of genome-to-genome sequence comparison. *Stand Genomic Sci* 2010;2:117–34.
- [25] Woese CR, Kandler O, Wheelis ML. Towards a natural system of organisms: proposal for the domains *Archaea*, *Bacteria*, and *Eucarya*. *Proc Natl Acad Sci USA* 1990;87:4576–9.
- [26] Gibbons NE, Murray RGE. Proposals Concerning the Higher Taxa of Bacteria. *Int J Syst Bacteriol* 1978;28:1–6.
- [27] Murray RGE. The Higher Taxa, or, a Place for Everything...? In: Holt JG, editor. *Bergey's Manual of Systematic Bacteriology*. First Edition Volume I. Baltimore, MD: The Williams and Wilkins Co.; 1984. p. 31–4.
- [28] Euzéby J. List of new names and new combinations previously effectively, but not validly, published. List no. 132. *Int J Syst Evol Microbiol* 2010;60:469–72.
- [29] Ludwig W, Schleifer KH, Whitman WB. Class I. *Bacilli* class nov. In: De Vos P, Garrity G, Jones D, Krieg NR, Ludwig W, Rainey FA, Schleifer KH, Whitman WB, editors. *Bergey's Manual of Systematic Bacteriology*. Second Edition Volume 3. New York: Springer-Verlag; 2009. p. 19–20.
- [30] Ludwig W, Schleifer KH, Whitman WB. Order II. *Lactobacillales* ord. nov. In: De Vos P, Garrity G, Jones D, Krieg NR, Ludwig W, Rainey FA, Schleifer KH, Whitman WB, editors. *Bergey's Manual of Systematic*

- Bacteriology. Second Edition Volume 3. New York: Springer-Verlag; 2009. p. 464.
- [31] Skerman VBD, McGowan V, Sneath PHA. Approved Lists of Bacterial Names. *Int J Syst Bacteriol* 1980;30:225–420.
- [32] Deibel RH, Seeley HW. Family II. *Streptococcaceae*. In: Buchanan RE, Gibbons NE, editors. *Bergey's Manual of Determinative Bacteriology*. Eighth Edition. Baltimore, MD: The Williams and Wilkins Co.; 1974. p. 490–515.
- [33] Rosenbach FJ. In: Bergmann JF (ed), *Microorganismen bei den Wund-Infektions-Krankheiten des Menschen.*, Wiesbaden; 1884, p. 1–122.
- [34] Deibel RH, Seeley HW. Genus I. *Streptococcus* Rosenbach 1884, 22. In: Buchanan RE, Gibbons NE, editors. *Bergey's Manual of Determinative Bacteriology*. Eighth Edition. Baltimore: The Williams and Wilkins Co.; 1974. p. 490–509.